IN THE CLAIMS:

1. (currently amended) A video conferencing system comprising:

a stationary image pickup device, remaining motionless during operation, for generating image signals representative of an image;

an audio pickup device for generating audio signals representative of sound from an audio source; and

~~a multimodal integration architecture system adapted~~ means for processing said image signals and said audio signals to determine a direction of the audio source relative to a reference point~~, the system being adapted to determine the direction~~ depending at least at times on the image signals.

2. (original) The video conferencing system of claim 1 wherein said ~~multimodal integration architecture system further~~ processing means comprises:

an audio source localization system;

a computer vision person detection system; and

a multimodal speaker detection system.

3. (original) The video conferencing system of claim 2, further comprising an integrated housing for an integrated video conferencing system incorporating the image pickup device, the audio pickup device, and the multimodal integration architecture system.

4. (original) The video conferencing system of claim 3, wherein the integrated housing is sized for being portable.

5. (previously presented) The video conferencing system of claim 1, further comprising an electronic pan tilt zoom system for electronically manipulating the image signals to effectively provide at least one of variable pan, tilt, and zoom functions.

6. (currently amended) The video conferencing system of claim 1, wherein the image pickup device is a stationary camera that remains motionless during operation of the video conferencing system.

7. (currently amended) The video conferencing system of claim 1, wherein ~~the multimodal integrated architecture system~~ processing means provides control signals to an electronic pan tilt zoom system.

8. (previously presented) The video conferencing system of claim 2, wherein the audio source localization system detects the movement of the audio source when the audio source moves relative to the reference point, and, in response to the movement, the audio source localization system causes a change in a field of view of the image pickup device.

9. (previously presented) The video conferencing system of claim 1, wherein the audio pickup device is comprised of an array of two microphones.

10. (previously presented) A method comprising the steps of:

generating, at a stationary image pickup device, remaining motionless during operation, image signals representative of an image;

generating, at an audio pickup device, audio signals representative of sound from an audio source;

processing the image signals and the audio signals to determine a direction of the audio source relative to a reference point, the determination depending at least at times on the image signals;

manipulating the image signals to produce refined image signals depending on the determined direction; and

outputting said refined image signals.

11. (original) The method of claim 10 further comprising the steps of:

applying said audio signals to an audio source localization system;

applying said image signals to a computer vision person detection system;

processing said audio signals and said image signals with a multimodal speaker detection system to determine the direction of the audio source;

generating control signals based on the determined direction of the audio source, the determination depending at least at times on the image signals;

applying the control signals to an electronic pan tilt zoom system to mimic the effect of at least one function of a movable camera, said function selected from the group consisting panning, tilting, and zooming said movable camera; and

providing an output from said electronic pan tilt zoom system.

12. (original) The method of claim 10, wherein manipulating the image signals includes varying a field of view of the image pickup device in response to the control signals.

13. (original) The method of claim 10, wherein processing the audio signals includes determining an audio based direction of the audio source based on the audio signals.

14. (previously presented) The method of claim 10, wherein processing the audio signals includes detecting the movement of the audio source when the audio source moves; and

manipulating the image signals includes causing electronically, in response to the movement, a variation in a field of view of the image pickup device.

15. (original) The method of claim 13, wherein processing the image signals includes generating control signals depending on the audio based direction, and manipulating the image includes electronically panning, tilting, and/or zooming said image pickup device depending on the control signals.

16. (previously presented) A video conferencing system comprising:

microphones for generating audio signals representative of sound from a speaker;

a stationary video camera, remaining motionless during operation, for generating video signals representative of a video image;

an electronic pan tilt zoom system for manipulating video images to produce the visual effects of panning, tilting, and/or zooming;

a processor for processing the video signals and the audio signals to determine a direction of a speaker relative to a reference point and supplying control signals to the electronic pan tilt zoom system for producing images that include the speaker in the field of view of the camera, the determination of direction depending at least at times on the video signals, the control signals being generated based on the determined direction of the speaker; and

a transmitter for transmitting audio and video signals for video conferencing.

17. (previously presented) The video conferencing system of claim 1, wherein at times the determination of the direction of the audio source depends on both the image signals and the audio signals.

18. (previously presented) The video conferencing system of claim 1, wherein the processing includes determining the movement of the audio source depending at least at times on the image signals.

19. (previously presented) The video conferencing system of claim 1, wherein the processing includes tracking the position of the audio source when the audio source moves, the tracking depending at least at times on the image signals.

20. (previously presented) The video conferencing system of claim 2, wherein the computer vision person detection system detects the movement of the audio source when the audio source moves relative to the reference point, and, in response to the movement, the computer vision

person detection system causes a change in a field of view of the image pickup device.

21. (previously presented) The method of claim 10, wherein processing the image signals further includes:

detecting the movement of the audio source when the audio source moves; and

causing electronically, in response to the movement, an variation in a field of view of the image pickup device.

R126
22 ~~23~~. (previously presented) The method of claim 10, wherein the processing includes determining the movement of the audio source depending at least at times on the image signals.

23 ~~24~~. (previously presented) The method of claim 10, wherein the processing includes tracking the position of the audio source when the audio source moves, the tracking depending at least at times on the image signals.

24 ~~25~~. (previously presented) A video conferencing system, comprising:

a stationary image pickup device, remaining motionless during operation, for generating image signals representative of an image;

an audio pickup device for generating audio signals representative of sound from an audio source;

means for processing the image signals and the audio signals to determine a direction of the audio source relative to a reference point, the determination depending at least at times on the image signals;

means for manipulating the image signals to produce refined image signals depending on the determined direction; and

an output for outputting said refined image signals.

25 ~~26~~. (new) The video conferencing system of claim 9, wherein the array of microphones includes only two microphones.